Préservation des métriques de confiance

Mots clefs

— Apprentissage Profond : l'apprentissage profond est ici restreint à l'entraînement et la définition de réseaux de neurones artificiels.

— Compression : la compression en apprentissage profond correspond à l'objectif de réduire le temps d'inférence, la consommation électrique ou l'empreinte mémoire d'un réseau de neurones artificiels.

— Confiance : dans le cadre de la compression, la confiance correspond à la capacité de la méthode de compression à préserver les propriétés mathématiques de la fonction de prédiction initiale.

— Lipschitz : Un fonction est dite Lipschitz lorsqu'il existe une constante C telle que pour tout x, y vérifiant $\|x-y\| < r$ on a $\|f(x) - f(y)\| < Cr$. Autrement dit, une fonction Lipschitz présente une robustesse de l'ordre de C aux petites variations sur ses entrées.

Contexte

La compression et l'accélération des réseaux de neurones artificiels est devenu un point central dans les produits de Datakalab et constitue un élément majeur dans le déploiement à l'échelle des solutions de vision par ordinateur. La compression des réseaux de neurones artificiels peut s'opérer de nombreuses façons dont l'élagage (pruning) et la quantification (quantization). L'équipe de recherche de Datakalab se concentre sur ces sujets avec pour objectif de proposer les réseaux les plus légers en terme de consommation mémoire et énergétique tout en préservant la précision des modèles sur les jeux de données de tests connus et standards dans le domaine [1,2,3]. Cependant, ces méthodes ne se font pas sans perte par rapport au réseau initial. L'objectif de cette thèse est d'une part, d'étudier et établir les conséquences théoriques et pratiques de la compression sur les fonctions de prédiction en terme d'écart au réseau initial et, d'autre part, de proposer des méthodes pour y remédier.

Objectifs

La préservation de la précision du modèle est une première métrique naïve permettant d'évaluer la qualité de compression. Cependant, dans le cas de systèmes critiques, cette métrique n'est pas suffisante. Un premier travail consistera alors à établir des mesures permettant de déterminer si une méthode de compression opère un changement conséquent ou non sur la fonction de prédiction. Dans un second temps, le travail se portera sur la conception et la mise en pratique de méthodes permettant de compresser les réseaux de neurones artificiels de façon robuste aux précédentes métriques établies.

Il existe actuellement de nombreux résultats sur le fonctionnement des réseaux de neurones, comme des propriétés d'approximation universelle mais également des garanties de convergence sous certaines contraintes. Nous avons également des résultats sur les propriétés des fonctions apprises par les réseaux de neurones artificiels. Entre autres, les réseaux contenant des fonctions d'activations ReLU définissent des fonctions affines par morceaux. Malgré cela, ces fonctions peuvent admettre des constantes de Lipschitz très grandes [4] et donc être très sensibles à de faibles variations sur les entrées. Il existe des travaux ayant pour but de réduire la constante de Lipschitz associée au réseau entraîné afin d'obtenir une plus grande robustesse lors de prédiction [5]. Cependant, il n'existe pas encore de travaux sur l'influence des méthodes de compression sur ces propriétés. Le premier objectif de cette thèse est donc d'établir la relation entre les méthodes de compression et les constantes de Lipschitz des réseaux entraînés.

Il existe également d'autres façons de mesurer la robustesse des réseaux aux attaques comme par exemple les méthodes adversaires [6]. Ces méthodes sont beaucoup plus étudiées que les approches théoriques et ont déjà vu de nombreux résultats intégrant l'influence de la compression [7]. Cependant la littérature actuelle n'a pas encore convergé à un consensus et les méthodes d'attaques contradictoires ne sont pas nécessairement réalistes. Un des objectifs de la thèse consistera à adapter et tester des attaques contradictoires au contexte de la compression plutôt que d'appliquer la compression en réponse à une attaque adversaire.

Enfin, l'explicabilité des réseaux est un élément incontournable de la confiance en apprentissage profond. Un volet conséquent de cette thèse portera sur la préservation de l'attribution [8]. En effet, dans de nombreux domaines d'applications l'attribution permet de confirmer l'utilisation des informations adéquates lors de la prise de décision. Cependant, il n'existe pas de garanties théoriques et peu de garanties pratiques sur la capacité des méthodes de compression à préserver l'attribution.

Après avoir établi les points cruciaux à la robustesse des réseaux compressés dans le cas d'une tâche sur un système critique, l'objectif sera de développer des méthodes de compression permettant d'améliorer la dite robustesse tout en maintenant les performances de compression [9].

Profil et compétences recherchées

Diplôme de Master ou Grande École. Compétences requises :

— Machine Learning / Deep Learning
— Vision par ordinateur
— Programmation Python et librairie deep learning (tensorflow ou pytorch)
— Excellentes capacités relationnelles et rédactionnelles (français et anglais)

Modalités de candidature

Pour postuler à cette thèse, le candidat est invité à communiquer par mail à kb@datakalab.com et lf@datakalab.com :

— Le CV
— Les résultats académiques des deux dernières années universitaires
— Un lien vers un projet de machine learning (lien GitHub / GitLab ou Colab)

Environnement

La thèse se déroulera dans le cadre d'un contrat de collaboration CIFRE entre la société Datakalab et l'Institut des Systèmes Intelligents et de Robotique (ISIR) de Sorbonne Université.

Datakalab est une startup spécialisée dans des algorithmes d'apprentissage profond à faible consommation, efficaces en termes d'exécution, respectueux de la vie privée et fonctionnant entièrement en embarqué. Ses travaux de recherches ont données lieux à des publications dans les meilleures conférences et journaux du domaine (T-PAMI, NeurIPS, ICCV, CVPR, AAAI)

L'**ISIR** est une unité mixte de recherche sous la tutelle de Sorbonne Université, du Centre National de la Recherche Scientifique (CNRS), et de l'Inserm, dont la recherche pluridisciplinaire rassemble des chercheuses, chercheurs, enseignantes-chercheuses et enseignants-chercheurs relevant de différentes disciplines en robotique, apprentissage machine, sciences du vivant et sciences médicales.

La thèse sera encadrée par Kévin Bailly directeur de la recherche de Datakalab et Maître de conférences, HDR, à l'ISIR et Arnaud Dapogny chercheur en IA à Datakalab.

Références

[1] *RED : Looking for Redundancies for Data-FreeStructured Compression of Deep Neural Networks*, 2021, NeurIPS, Yvinec Edouard, Dapogny Arnaud, Cord Matthieu and Bailly, Kévin
[2] *RED++ : Data-Free Pruning of Deep Neural Networks via Input Splitting and Output Merging*, 2022, TPAMI, Yvinec Edouard, Dapogny Arnaud, Cord Matthieu and Bailly, Kévin
[3] *To Fold or Not to Fold : a Necessary and Sufficient Condition on Batch-Normalization Layers Folding*, 2022, IJCAI, Yvinec Edouard, Dapogny Arnaud, Cord Matthieu and Bailly, Kévin
[4] *Efficient and accurate estimation of lipschitz constants for deep neural networks*, 2019, NeurIPS, by Fazlyab Mahyar, Robey Alexander, Hassani Hamed, Morari Manfred and Pappas George
[5] *Lipschitz-margin training : Scalable certification of perturbation invariance for deep neural network*, 2018, NeurIPS, by Tsuzuku Yusuke, Sato Issei and Sugiyama Masashi
[6] *Threat of adversarial attacks on deep learning in computer vision : A survey*, 2018, IEEE access, by Akhtar Naveed and Mian Ajmal
[7] *Qusecnets : Quantization-based defense mechanism for securing deep neural network against adversarial attacks*, 2019, IEEE access, by Khalid Faiq and Ali Hassan and Tariq Hammad and Hanif Muhammad Abdullah and Rehman Semeen and Ahmed Rehan and Shafique Muhammad
[8] *Improving performance of deep learning models with axiomatic attribution priors and expected gradients*, 2021, Nature machine intelligenc, by Erion Gabriel, Janizek Joseph D, Sturmfels Pascal, Lundberg Scott M and Lee Su-In
[9] *Defensive Quantization : When Efficiency Meets Robustness*, 2018, ICML, by Lin Ji, Gan Chuang and Han Song